Identification of potential molecular associations between chikungunya virus non-structural protein 2 and human host proteins

J. RANA1*, S. GULATI1*, S. RAJASEKHARAN1, A. GUPTA4, V. CHAUDHARY5, S. GUPTA1

1Center for Emerging Diseases, Department of Biotechnology, Jaypee Institute of Information 4 Technology, Noida, U.P, India-201307; 2Department of Pharmacology, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA; 3Division of Molecular Diagnostics of Oncogenic Infections (F020), German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; 4Department of Microbiology, University of Delhi, Benito Juarez Marg, New Delhi 110021, India; 5Department of Biochemistry, University of Delhi, Benito Juarez Marg, New Delhi 110021, India

Received March 8, 2016; revised July 17, 2016; accepted January 6, 2017

Summary. – Chikungunya virus (CHIKV) non-structural protein 2 (nsP2) is considered to be the master regulator of viral RNA replication and host responses generated during viral infection. This protein has two main functional domains: an N-terminal domain which exhibits NTPase, RNA triphosphatase and helicase activities and a C-terminal protease domain. Understanding how CHIKV nsP2 interacts with its host proteins is essential for elucidating all the required processes for viral replication and pathogenesis along with the identification of potential targets for antiviral therapy. In current study yeast two-hybrid (Y2H) screening of a human fetal brain cDNA library was performed using nsP2 protein as bait. The analysis identified seven host proteins (CCDC130, CPNE6, POLR2C, MAPK9, EIF4A2, EEF1A1 and EIF3I) as putative interactors of CHIKV nsP2 which were selected for further analysis based on their roles in host cellular machinery. The gene ontology analysis indicates that these proteins are mainly involved in apoptosis, transcription and translational mechanism of host cell. Domain mapping of nsP2 revealed that these associations are not random connections but instead they have functional significance. Further studies to identify the amino acid residues and their chemical interactions that may help in opening new possibilities for preventing these interactions, thus reducing chances of chikungunya infection were performed. This study expands the understanding of CHIKV-host interactions and is important for rational approaches of discovering new antiviral agents.

Keywords: chikungunya virus; non-structural protein 2; host-virus interactome; yeast two-hybrid; pull down; ELISA

Introduction

Chikungunya virus (CHIKV; the family Togaviridae, the genus Alphavirus) is an arthritogenic old world alphavirus causing epidemics in the tropical regions of Asia and Africa, and is also spreading quickly to temperate climates with recent documented outbreaks in Europe and America. The re-emergence of CHIKV with unprecedented magnitude and unusual clinical severity involving neurotropism has posed a major public health concern worldwide, with no preventive or therapeutic means available. Altogether, this demands the need for the development of vaccine and therapeutic strategies for prevention of human population from this virus. Although CHIKV associated research has been initiated a decade ago, the availability of CHIKV specific reagents for in-depth investigation of viral infection and replication is scanty.

Among the nine proteins encoded by CHIKV genome, non-structural protein 2 (nsP2) is a multifunctional protein,
known to play key regulatory roles in both viral replication and host responses (Vasiljeva et al., 2009; Karpe et al., 2011; Breakwell et al., 2007; Gorbalenya and Koonin, 1993). The N-terminal domain exhibits NTPase, helicase and RNA triphosphatase activities (Karpe et al., 2011; Gorbalenya and Koonin, 1993) and C terminal domain possess a cysteine protease required for cleavage of nonstructural polyprotein complex (Vasiljeva et al., 2009). Moreover, the alphavirus nsP2 protein has been described as the virulence factor responsible for the transcriptional and translational shut-off in infected host cells, and the inhibition of interferon mediated antiviral responses contributing to the controlling of translational machinery by viral factors (Breakwell et al., 2007). These functions performed by nsP2 during viral life cycle within susceptible host can be accomplished through association with a number of host factors (Bourai et al., 2012). Hence, targeting the residues within active site and those involved in host associations would be the most rationale strategy to inhibit its functions and consequently inhibiting viral replication.

The current study identified a number of cellular interactors of CHIKV nsP2 from human fetal brain cDNA library using yeast two-hybrid (Y2H) analysis. The identified interactions were confirmed by two independent in vitro assays viz. pull down and ELISA, followed by domain mapping of nsP2 to identify the specific domain involved in interactions. Further, we identified the amino acid residues that come in direct contact during these associations using protein docking studies. This study together with the knowledge of intraviral protein interaction (Sreejith et al., 2012; Rana et al., 2014) can provide molecular targets for the development of therapeutic strategies based on blockage of these associations.

**Materials and Methods**

**Screening of human fetal brain cDNA library for CHIKV nsP2 interactors using yeast two-hybrid system.** Mate & Plate-human fetal brain cDNA library (Clontech, USA) was used for Y2H screening to identify the host interactors of CHIKV nsP2 protein. The viral protein nsP2, cloned in pGBK7 (BD) yeast expression plasmid, was transformed into Saccharomyces cerevisiae AH109 strain (generated in our earlier work; Sreejith et al., 2012) and allowed to mate with S. cerevisiae Y187 strain pre-transformed with the human fetal brain cDNA library using protocols described in our previous works (Dudha et al., 2014; Rajasekharan et al., 2015). The mated diploid clones harboring both bait (BD, coding nsP2) and prey (library plasmids/pGADT7-Rec/AD; host proteins) plasmids were initially selected on triple (TDO; SD-Trp/-Leu/-His – synthetic dropout media lacking amino acids tryptophan, leucine and histidine, and the nitrogen base adenine) supplemented with X-a-Gal. Following this, multiple prey plasmids were eliminated by blue-white screening and duplicate clones using Haet III restriction digestion analysis. The 5’ and 3’ ends of human cDNA within each selected prey plasmid were determined by sequencing using T7 forward (5’-TAATACGACTCACTATAGGGG-3’) and AD specific reverse primers (5’-AGATGGTGCCAGCAGGCACAG-3’). The sequence data were used in BLAST suite of tools to query the NCBI human genome and human transcript databases.

**Pull down and ELISA.** The interacting host genes were cloned in bacterial expression vectors (pGEX-4T3 or pCAK) using CDS of host protein in pGADT7-Rec plasmids (library/prey plasmids) as templates to amplify the corresponding host nucleotide sequence and gene specific primers flanked with appropriate restriction sites at their 5’ end. The host constructs together with the viral nsP2 and its functional domains constructs were individually transformed and over-expressed in Escherichia coli BL21 cells using 0.5% arabino (pCAK vector) or 1 mmol/l isopropyl β-D-1-thiogalactopyranoside (IPTG; pGEX-4T3 vector) for 4 h at 25°C (Rana et al., 2014; Sreejith et al., 2012). Following induction, the induced cells were harvested and lysed using lysis buffer (IBA-GmbH, Germany). The lysates (soluble protein fraction) were used to perform pull down and protein interaction ELISA using protocols described in our earlier works (Sreejith et al., 2012; Kumar et al., 2011, 2012).

**Identification of interacting residues.** The structures of CPNE6 and CCDC130 were generated using I-TASSER (Zhang, 2008). During the comparative modeling procedure, the homology models of EEF1A1, EIF3I, EIF4A2, MAPK9 and POLR2C were constructed with MODELLER ver. 9.4 (Sali et al., 1995) using templates provided in Supplementary file 1 as described earlier (Sreejith et al., 2013). Briefly, pairwise sequence alignments between the query proteins and the template sequences were carried out using ClustalW. BLOSUM62 was used as the scoring matrix and default values were used for all other parameters. The resulting alignments were given as inputs to MODELLER ver. 9.4 to build protein models with a high degree of optimization level with simulated annealing. The loops of protein models were further refined individually using MODELLER’s loop optimization procedure with high resolution discrete optimized protein energy (DOPE) method. Due to the unavailability of a full-length structural template for CHIKV nsP2 protein, the structures of CHIKV nsP2 helicase and protease domains were generated using I-TASSER and MODELLER ver. 9.4, respectively, followed by fusing these domains to reconstitute the full-length CHIKV nsP2 structure. The atomic clashes and bond length errors within these structures were repaired by using MoRefiner (Xu and Zhang, 2011). The generated protein structures were energy minimized and protonated before predicting their secondary structure using I-TASSER (Zhang, 2008). The structures of CHIKV nsP2 helicase and protease domains were generated using I-TASSER and MODELLER ver. 9.4, respectively. The generated protein structures were energy minimized and protonated before predicting their secondary structure using I-TASSER (Zhang, 2008).
approach as described earlier (Rana et al., 2014). Briefly, biological data from the experiments were used to either prevent undesirable contacts within ZDOCK run or to select for known contacts by using distance filters on predictions after RDOCK refinement (Li et al., 2003). The following steps involved re-ranking and clustering ZDOCK’s top 2000 initial predictions with ZRANK (Pierce and Weng, 2007), followed by refining the top 100 cluster centers with RDOCK. Poses belonging to the cluster containing the top cluster center were refined and then the lowest energy members of the cluster were visually examined and selected. The protein molecules were typed with and corrected for CHARMm Polar H force field prior to RDOCK refinement. Lastly, the interface residues were identified with LIGPLOT (Wallace et. al., 1996).

Results

Screening of fetal brain cDNA library for nsP2 interactors

CHIKV nsP2 cloned in pGBKT7 vector was transformed in to AH109 cells and mated with fetal brain cDNA library pretransformed in Y187 yeast cells. The screening of diploid cells, initially on medium stringency media followed by selection under high stringency conditions, yielded around 200 clones as nsP2 interactors. The PCR analysis of clones obtained from QDO (SD-Trp/-Leu/-His/-Ade) media using AD specific primers showed that insert size ranged from 500 bp to 4 kb. The PCR analysis showed that some of the clones contained multiple plasmids and a number of different clones contained inserts of same sizes or types. The clones with multiple bands were selected on QDO media supplemented with X-α-Gal to segregate the plasmids and only blue colonies were selected for further studies as these represent the true interactors while white colonies were discarded. The HaeIII restriction digestion profiling of clones with same insert sizes revealed that most of these clones represent a different insert. Among 150 unique clones obtained after removal of multiple and duplicate plasmids, 45 clones were selected for further analysis based on their growth rate on QDO media. These clones produced colonies within 6 days after streaking on QDO supplemented with X-α-gal media and hence considered as strong interactors of nsP2. The interactions were confirmed by more than three rounds of testing in yeast under high stringency conditions to eliminate false positives.

The sequencing analysis of prey plasmid isolated from selected clones identified 27 human gene sequences while rest of the sequences involved chromosome contigs and uncharacterized loci. Subsequent analysis of amino acid sequences of all the identified clones using ExPASy showed that among 27 identified sequences only 17 clones represent either complete ORF or a region of known human proteins (Table 1). This could either be due to the incorrect reading frame or the cDNA insert is from the region outside the CDS of the gene.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Clone No.</th>
<th>Proteins identified</th>
<th>Size of library clone (aa)</th>
<th>Size of human protein (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>Copine VI (CPNE6)</td>
<td>558</td>
<td>558</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>Eukaryotic translation initiation factor 4A2 (EIF4A2)</td>
<td>85</td>
<td>408</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>Eukaryotic translation elongation factor 1 (EEF1A1)</td>
<td>236</td>
<td>463</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>Cytochrome C oxidase subunit VIIa polypeptide 2 (COX7A2)</td>
<td>71</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>Coiled coil domain containing 130 (CCDC130)</td>
<td>397</td>
<td>397</td>
</tr>
<tr>
<td>6</td>
<td>81</td>
<td>RNA polymerase II (DNA directed) polypeptide C (POLR2C)</td>
<td>296</td>
<td>296</td>
</tr>
<tr>
<td>7</td>
<td>85</td>
<td>Nardilysin (N-Arginine dibasic convertase) (NRD1)</td>
<td>85</td>
<td>1087</td>
</tr>
<tr>
<td>8</td>
<td>94</td>
<td>OUT domain, ubiquitine aldehyde binding 1 (OTUB1)</td>
<td>78</td>
<td>272</td>
</tr>
<tr>
<td>9</td>
<td>95</td>
<td>NSFL1 (p97) cofactor (p57) (NSFL1C)</td>
<td>371</td>
<td>371</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>Actin gamma 1 (ACTG1)</td>
<td>196</td>
<td>375</td>
</tr>
<tr>
<td>11</td>
<td>100</td>
<td>Eukaryotic translation initiation factor 3 Subunit I (EIF3I)</td>
<td>184</td>
<td>326</td>
</tr>
<tr>
<td>12</td>
<td>108</td>
<td>Mitogen activated protein kinase 9 (MAPK9)</td>
<td>61</td>
<td>425</td>
</tr>
<tr>
<td>13</td>
<td>209</td>
<td>Mitochondrial translational release factor 1 (MTRF1L)</td>
<td>142</td>
<td>272</td>
</tr>
<tr>
<td>14</td>
<td>210</td>
<td>Mannosidase, alpha, class 2a, member 2 (MAN2A2)</td>
<td>595</td>
<td>1151</td>
</tr>
<tr>
<td>15</td>
<td>267</td>
<td>Flotilin 1 (FLOT1)</td>
<td>226</td>
<td>428</td>
</tr>
<tr>
<td>16</td>
<td>293</td>
<td>Copper metabolism (murr1) domain containing 1 (COMMD1)</td>
<td>191</td>
<td>191</td>
</tr>
<tr>
<td>17</td>
<td>326</td>
<td>Tyrosine phosphate receptor F (PTPRF)</td>
<td>108</td>
<td>1899</td>
</tr>
</tbody>
</table>

The gene sequences of the identified clones were translated into protein using ExPASy translation tool. The protein sequences obtained were analyzed using protein Blast (pBlast). This analysis showed that only 17 proteins aligned with human proteins representing complete or partial CDS of the respective protein.
Fig. 1

Interaction analysis among nsP2 and host proteins using pull down

Western blotting of human host proteins interacting with nsP2. The presence of fusion proteins in eluates was detected using both anti-Strep and anti-GST antibodies. (a) The presence of Strep-CPNE6 and GST-nsP2 proteins (lane 1); Strep-nsP2 and GST-CCDC130 (lane 2); Strep-nsP2 and GST-MAPK9 (lane 3); Strep-nsP2 and GST-EIF4A2 (lane 4); Strep-nsP2 and GST-POLR2C (lane 5); Strep-nsP2 and GST-EIF1A1 (lane 6); Strep-nsP2 and GST-EIF3I (lane 7). (b) Western blot of nsP2 helicase domain D1 (nsPD1) interaction with CPNE6 (lane 1), EEF1A1 (lane 6) and EIF3I (lane 7). The absence of protein bands corresponding to GST-CCDC130 (lane 2; anti-GST), GST-MAPK9 (lane 3; anti-GST), GST-EIF4A2 (lane 4; anti-GST), GST-POLR2C (lane 5; anti-GST) and presence of Strep-nsP2D1 (lane 2–5; anti-Strep) shows that the helicase domain of nsP2 does not interact with these proteins. (c) Western blot of nsP2 protease domain D2 (nsPD2) interaction with CPNE6 (lane 1), CCDC130 (lane 2), MAPK9 (lane 3), EIF4A2 (lane 4) and POLR2C (lane 5). The absence of band corresponding to GST-EEF1A1 (lane 6; anti-GST) and GST-EIF3I (lane 7; anti-GST) while presence of Strep-nsP2D2 (lane 6 and 7; anti-Strep) showed that these proteins do not interact with nsPD2. M is the prestained protein marker (Fermentas).

Validation of Y2H screening and domain mapping of nsP2

Among the 17 host proteins identified by Y2H assay, seven proteins (CCDC130, CPNE6, EEF1A1, EIF4A2, EIF3I, MAPK9 and POLR2C) were selected for further confirmation based on two criteria. The first criteria is based on the observation that protein-protein interactions detected by multiple independent assays or studies for other viruses are
more likely to be biologically significant and second criteria is their role in host cellular machinery. The hub proteins CPNE6, MAPK9 and CCDC130 are involved in multiple signaling pathways of the host cells and were previously reported to play role in viral infection (Labrada et al., 2002; Park et al., 2001) and interaction with flavivirus nsPs (Le Breton et al., 2001). The proteins of host transcriptional and translation machinery POLR2C, EIF4A2, EIF3A and EEF1A1 identified as cellular interactor of nsP2 might be required for replication and translation of viral genome as the virus depends on host proteins for the production of viral proteins. Further, the specific domain of nsP2 involved in these interactions was identified. As nsP2 is composed of two functional domains: N-terminal helicase (1–456 aa) and C-terminal protease (456–798 aa), mapping of domains involved in the identified interactions will help in better understanding of their roles in viral pathogenesis. The association of nsP2-host proteins and domain mapping were performed using two independent assays; pull down and ELISA. The soluble fraction of recombinant proteins obtained after lysis of induced bacterial cultures was used for these assays. The strep fusion proteins were used as bait to capture GST tagged interacting partner. For pull down assay, the strep tagged fusion protein was allowed to bind on streptactin resin providing a secondary platform for the binding of GST tagged interacting partner. The immunoblotting analysis of eluted protein complexes bound to streptactin resin was performed with both anti-Strep and anti-GST antibodies. This assay confirmed the association of all the host proteins selected from Y2H screening.
with full length nsP2 as both fusion proteins were co-eluted as a complex from the resin (Fig 1a; lane 1–7). The analysis of eluates from domain-host test pairs revealed that CPNE6 interacted with both helicase and protease domain of nsP2 as indicated by the presence of both Strep and GST tagged proteins (Fig 1b,c; lane 1), EEF1A1 and EIF3I with helicase domain only and (Fig 1b; lane 6 and 7) not with protease domain (Fig 1c; lane 6 and 7); CCDC130, MAPK9, EIF4A2 and POLR2C did not interact with helicase domain (Fig. 1b; lane 2–5) instead were found to interact with protease domain of nsP2 (Fig 1c; lane 2–5).

Further, these results were confirmed by ELISA using streptactin coated microtitre plates. The OD_{450} values obtained, showed that all the host proteins selected were interactors of CHIKV nsP2. The domain mapping using protein interaction ELISA produced the same results as observed by pull down analysis. The experimental controls (only GST fusion proteins), positive (Chandipura virus M-N interaction) and negative (Chandipura virus M-P interaction) controls used during this experiment produced the expected results. All the experiments were performed in triplicates and the average of OD_{450} values with standard error have been represented graphically (Fig. 2). The OD_{450} value of 0.15 was taken as a cutoff for presence or absence of interaction. The test pairs having OD_{450} below 0.15 were taken as negative while above this were considered positive.

Identification of interacting residues

Following the generation of protein models using a combination of I-TASSER and MODELLER ver. 9.4, the interaction models of CHIKV nsP2 with individual host proteins were constructed using rigid body docking in combination with rescoring and refinement (Fig. 3). ZDOCK ver.2.3 was run on each of the following protein pairs: nsP2-CCDC130, nsP2-CPNE6, nsP2-EEF1A1, nsP2(EIF3I, nsP2(EIF4A2, nsP2-MAPK9 and nsP2-POLR2C to produce initial sets of structures for refinement. ZRANK was then used to re-rank the top 2000 poses out of the 3600 poses from the initial-stage docking predictions for each ZDOCK run. RDOCK was used to refine the top 100 re-ranked cluster centers to identify the best cluster center, whose corresponding cluster was later RDOCK refined to obtain its lowest energy member. All the protein pairs were found to be in agreement with the biological data from the experiments upon analyzing their corresponding binding interfaces. The complimentary residues involved, mainly interact through electrostatic, hydrogen and hydrophobic bonds (Supplementary file 2).

Discussion

The present study identifies the binding partners of CHIKV nsP2 from human fetal brain cDNA library us-
As CHIKV nsP2 is composed of two functionally distinct domains performing different functions during viral life cycle, domain mapping would help in deducing relevance of these interactions. The helicase domain of nsP2 was found to interact with EEF1A1 and EIF3I. EEF1A1 forms a trimeric complex with PARP1 and Txk, which is involved in Th1 cell cytokine production (Maruyama et al., 2013), and hepatocyt E virus (Geng et al., 2009). The interaction of nsP2 might inhibit the formation of this complex and thus the cytokine production. EIF3I is involved in multiple steps of translation initiation and interaction of helicase domain with this protein might be required for the translation initiation of newly synthesized positive strand viral RNA. The C terminal protease domain of nsP2 interacted with EIF4A2, CCDC130, POLR2C and MAPK9. These proteins are mainly involved in transcription-translation processes and apoptosis which are attributed to be affected by C terminus of nsP2 and hence enhance the reliability of interaction dataset obtained in this study. The host protein EIF4A2 is involved in cap recognition and required for binding of mRNA to the ribosome. As herpes simplex virus virion host shut-off protein is known to interact with this protein along with EIF4H which helps the virus in redirecting the cell from host to viral protein synthesis (Feng et al., 2005), interaction of EIF4A2 protein with CHIKV nsP2 might also be involved in enhancing viral protein generation and inhibiting host protein translation. POLR2C is a component of DNA dependent RNA polymerase II and nsP2 is known to cause degradation of POLR2A (RPB1; another subunit of this enzyme, Akhrymuk et al., 2012). Binding of POLR2C with nsP2 might help in locating this complex to the enzyme and enhance the degradation of RPB1 leading to inhibition of host protein synthesis and finally cell death. MAPK9 and CCDC130 are involved in the signaling pathway that leads to apoptosis; their interaction with nsP2 might be involved in apoptosis regulation during CHIKV infection.

Although, the functions of individual domains of nsP2 and host protein shows the significance of these interactions, further experimental work is required to study the biological relevance of these data. Also, the amino acid residues identified in this study might play crucial roles in maintaining these interactions between virus and host at molecular level and alteration of these residues may change the interacting complementary surfaces thus destabilizing and/or preventing formation of these complexes. In conclusion, the identification of cellular partners for nsP2 is expected to provide a starting point for further search of possibilities for therapeutic interventions to suppress CHIKV infection.

Acknowledgement. This work was funded by research grant from Department of Biotechnology, Government of India (Grant No. BT/PR11162/MED/29/97/2008).

Supplementary information is available in the online version of the paper.

References


